Multimodalité et sémantique.

José Rouillard

Multimodalité et sémantique.

José Rouillard

Chapitre 8

Multimodalité et sémantique

José Rouillard

Résumé

Ce chapitre traite de la sémantique et de la multimodalité dans le domaine des interactions Homme-Machine. Nous exposons un bref rappel des notions de média, multimédia, mode, modalité et multimodalité, avant de montrer en quoi la sémantique est une notion essentielle qui différencie un système informatique multimédia d'un système informatique multimodal. Du fait que plusieurs approches sont possibles en conception d'interfaces Homme-Machine, selon que l'on se place côté machine, côté utilisateur, on encore de manière hybride (machine/ utilisateur), la multimodalité s'inspire de modèles pluridisciplinaires et bénéficie des avancées en informatique, sociologie, psychologie, économie et théorie des jeux, voire des théories du marketing (le multicanal étant effectivement étudié lors de communication personne/organisation).

Nous illustrons notre propos grâce à la présentation d'un système de dialogue Homme-Machine orienté apprentissage, dans lequel un internaute interagit avec un agent animé. L'apport et l'enjeu de la multimodalité dans ce contexte est expliqué et commenté : l'agent est perçu visuellement, puisqu'il se déplace sur l'écran, mais aussi auditivement, car une voix de synthèse accompagne ses actions, et apporte des indications complémentaires ou redondantes à l'usager. Cette multimodalité véhicule également des éléments supplémentaires pour l'acquisition de sens : geste de désignation, comportements anthropomorphiques (humeur, émotions, aides personnalisées), animations permettant de comprendre ce qu'il se passe au niveau du noyau fonctionnel du système informatique, etc.

8.1 Les interactions Hommes-Machine

La majeure partie des tâches que nous confions à nos ordinateurs, n'est pas d'ordre purement calculatoire, comme on le pense souvent, mais plutôt d'ordre computationnel. En effet, le calcul, au sens mathématique du terme, n'est qu'une partie des traitements qu'effectuent nos machines. Quotidiennement, nous les sollicitions pour des tâches répétitives (traitement de texte, envoi de courrier électronique, recherche d'information, etc.), pour lesquelles nous manipulons des symboles. La plupart du temps, l'ordinateur n'est pas le destinataire de cette information, mais simplement un médiateur. Nous nous trouvons donc dans une situation de dialogue entre humains, instrumenté grâce à une chaîne d'outils.

Parfois, la machine peut devenir un interlocuteur, capable de "comprendre" certaines commandes ou certains éléments d'un dialogue. Dans ce cas, l'ordinateur n'est pas simplement utilisé pour transmettre des informations d'un humain vers un autre. Il doit les traiter en y donnant du sens. D'ailleurs, pour évaluer les capacités d'une machine, une possibilité – parmi d'autres – consiste à tester non pas seulement la puissance calculatoire de celle-ci, mais bien plus encore, ses facultés à entretenir une conversation avec un interlocuteur humain (voir à ce propos le fameux « test de Turing »). Le problème le plus délicat, que d'autres avant nous ont déjà mis en lumière, c'est sans doute que l'homme veut tenter de transmettre quelque chose (une intelligence ?) à la machine, alors même qu'il ne maîtrise pas complètement la nature et les phénomènes qui régissent son fonctionnement naturel. Dans son « Histoire universelle des chiffres », Georges Ifrah livre sa réflexion sur ce sujet, avec un paragraphe qu'il intitule « Lorsque la métaphore fut identifiée à la réalité », et dont voici un large extrait :

> « Dans leur enthousiasme, ces savants crurent d'abord que toute opération de la pensée humaine était exclusivement de nature calculatoire (c'est-à-dire algorithmique), et donc que tout processus intellectuel était exécutable sur une machine de type ordinateur.
>
> Et comme ils désignèrent sous le nom de neurones et de synapses des modules qui n'eurent en réalité qu'une très lointaine parenté avec les composants réels, beaucoup plus complexes, d'un encéphale vivant, ils établirent inévitablement, selon un anthropomorphisme simpliste, et forcément réducteur, un parallèle, très étroit entre les circuits d'un calculateur électronique avec les cellules nerveuses et les composants neurobiologiques du cerveau vivant.
>
> Et c'est ainsi que l'organe central d'un simple ordinateur électronique fut désormais considéré comme l'équivalent du cerveau humain, le public assimilant dès lors les activités d'un calculateur à celles, éminemment supérieures, de l'esprit humain.

D'où, par identification abusive de la métaphore avec la réalité, et par une sorte de convergence avec le mythe du « cerveau électronique », le développement de l'idée d'une machine prétendument douée de pensée inductrice, et créatrice, munie de la faculté de prendre toutes sortes d'initiatives, et à laquelle l'homme pourrait confier tous ses problèmes sans exception pour obtenir d'elle toutes les solutions voulues quasi instantanément.» [IFR 94].

On a donc tendance à faire, de manière trop hâtive, un parallèle entre les possibilités d'une machine et celles de l'homme, avec la fausse idée qu'une grande puissance calculatoire puisse être suffisante pour compenser de faibles capacités à donner du sens aux informations traitées.

8.2 Qu'est ce que computer ?

Depuis quelques décennies, nous sommes passés de systèmes technocentrés à des systèmes plus anthropocentrés, pour lesquels l'utilisateur est au cœur du processus d'interaction. Or, par nature, l'homme a la faculté de computer[1], parfois sans même savoir quels sont les mécanismes en jeu dans ses raisonnements. L'esprit humain est vu, par certains chercheurs, comme un système de manipulation de symboles (H. Simon), ou comme un système de traitement de l'information (A. Newell). Jean-Louis Le Moigne explique « *Computer, ce n'est pas seulement calculer arithmétiquement des nombres, c'est, très généralement, manipuler et traiter des symboles.* » [LEM 86].

Le contexte est très souvent un facteur permettant de désambiguïser un énoncé. Par exemple, une personne qui voit un panneau mentionnant « St. John St. » dans une ville anglaise n'aura guère de mal à traduire cela en « Saint John Street ». Henriette Walter montre que l'aspect géographique et territorial influence grandement le sens des mots et des expressions d'une langue [WAL 98]. Elle prend l'exemple de conversation suivant :

– remettez-vous donc !

– non merci, je reste droit.

– mais si, j'insiste, vous allez resquiller et vous ruiner la jambe.

– ça vaut mieux que de partir de cinq en cinq.

Un dialogue apparemment incohérent, et pourtant parfaitement logique (lorsque l'on est de Sète) :

---

[1] À prononcer à la française : c'est ici le verbe français, et non le substantif anglais. Du latin computus «compte». Littéralement «computer» signifie calculer. Plus exactement, calculer la date des fêtes mobiles. Ainsi le «comput ecclésiastique» dresse le calendrier de la date de Pâques.

– asseyez-vous donc !

– non merci, je reste debout.

– mais si, j'insiste, vous allez glisser et vous casser la jambe.

– ça vaut mieux que de décliner lentement.

Il faut donc computer les informations pour y trouver du sens. Au niveau langagier, l'émetteur d'un message fait parfois l'hypothèse que ses interlocuteurs vont décoder le message de manière approprié. C'est ce que Violaine Prince appelle les principes d'économie et d'expansion [PRI 96]. Par exemple, lorsque l'on dit « Ne jetez rien dans cette poubelle » et que l'on souhaite faire comprendre à l'interlocuteur « Ne jetez rien d'important, ou de confidentiel, dans cette poubelle ». D'autres chercheurs appellent cela l'effet paillasson [CHA 02] : en effet, lorsque l'on vous demande de vous essuyer les pieds sur un paillasson, il est évident que l'on vous suggère d'essuyer les semelles de vos chaussures ; on ne souhaite pas vous voir retirer vos chaussures et vos chaussettes, pour frotter la plante de vos pieds sur le paillasson. Ces traitements d'informations sont quasiment instinctifs pour les humains. Il n'en va pas de même pour les ordinateurs.

Cependant, la communication humaine ne s'appuie généralement pas sur un seul canal de communication, mais bien sur plusieurs, et le plus souvent, de manière simultanée. Dans un dialogue, le geste accompagne fréquemment le discours. [MCN 92] a suggéré que le geste et la parole étaient générés par la même région du cerveau, et d'autres pensent que les premiers langages humains étaient basés sur des gesticulations (Cf. Zimmer 95, cité par [THO 96]). Beaucoup de classifications ont été utilisées pour décrire les types de gestes que l'on utilise lors de dialogues ou de discours [RIM 91], [POY 80]. La plupart sont dérivées des travaux de [EFR 41]. Globalement, on retrouve entre quatre et six classes de gestes différentes. Les plus étudiées et les plus utilisées pour la communication homme-machine sont les suivantes : gestes non-descriptifs, descriptifs (idéographiques), iconographiques, pantomimiques, déictiques, emblématiques (symboliques) et métaphoriques. Certains chercheurs parlent également d'une autre catégorie (en anglais « self adaptors », [MCN 92], [EKM 69]), pour référencer les actions « personnelles » comme se passer la main dans les cheveux, se gratter, etc. Les gestes, mimiques et attitudes du visage ont également beaucoup été étudiés en communication homme-homme et homme-machine [EKM 75]. Et, là encore, des taxonomie de ces gestes ont été proposées : on note par exemple les gestes symboliques, émotionnels, affectifs, conversationnels, ponctuateurs, régulateurs, manipulateurs, etc. [EKM 78].

Au regard de ces travaux de recherches et des différentes observations effectuées en laboratoire, il nous semble que si nous voulons doter nos machines de facultés à

interpréter nos actes et notre langage, nous devons pousser plus avant les travaux dans le domaine de la sémantique couplée à la multimodalité.

8.3 La multimodalité en IHM

Nous avons vu précédemment que les ordinateurs ne servent pas uniquement à calculer des opérations, mais qu'ils permettent également de traiter, stocker et indexer des informations de différentes natures. Le problème principal que nous rencontrons lors des processus de recherche d'informations n'est plus la disponibilité de celles-ci mais plutôt la capacité de sélection d'une information qui réponde à nos besoins [NIG 01]. De plus la quantité d'information numérique disponible ne fait que croître ; cette augmentation ne peut pas être maîtrisée uniquement par la croissance en puissance des calculateurs. Un projet à l'Université de Berkeley a estimé à un exa-octet[2] (1 million de tera-octets) la quantité de données générées annuellement de par le monde. Parmi ces données, 99,997 % sont disponibles sous forme numérique [KEI 01].

Dès lors, quelles sont les idées novatrices les plus prometteuses pour permettre l'accès à ces informations par un large public ? Les Interfaces Homme/Machine (IHM) utilisent de plus en plus de multimodalité. Le dialogue ne se fait plus seulement avec des canaux de communication traditionnels (clavier/souris en entrée, et écran en sortie), mais avec plusieurs comme par exemple des observations par caméra, gestes de désignation, objets à retour d'efforts, etc. [ROU 01]. L'homme souhaite communiquer de manière naturelle et intelligente avec les machines. Pour cela, les chercheurs dotent les ordinateurs de capteurs et d'effecteurs. Mais l'efficacité de l'interaction avec l'usager n'est pas seulement fonction du matériel et des données, il faut encore que le système « donne du sens » aux éléments qu'il manipule, or « un ordinateur multimédia n'est pas forcément multimodal » [COU 95]. La conception des Interfaces Homme-Machine (IHM) se centre principalement sur l'utilisateur et les usages. De ce fait, le domaine de recherche des IHM se situe à l'intersection de nombreuses autres disciplines, comme l'informatique, l'ergonomie, la linguistique, la psychologie sociale et cognitive, etc.

Historiquement, c'est en 1983 que Nicholas Negroponte fonde le MediaLab au MIT avec l'objectif d'étudier l'intégration de nouvelles technologies comme la synthèse de la parole, la vision par ordinateur, la synthèse d'image, les nouveaux dispositifs d'interaction comme le visiocasque et les transducteurs gestuels rétroactifs. En France, c'est en 1990 à Grenoble, qu'émerge le concept de

[2] Un tera-octet = 2 puissance 40 octets ; un exa-octet = 2 puissance 60 octets.

multimodalité en informatique. L'apport de la multimodalité va se concrétiser notamment à travers la robustesse, la flexibilité, l'adaptabilité[3] et l'adaptativité[4] des interactions.

8.3.1 Terminologie

Étant donné que l'IHM est pluridisciplinaire, chaque discipline amène une définition particulière des concepts manipulés. Nous savons qu'une définition claire et précise de chaque terme n'est pas toujours possible, et que cela donne lieu à des interprétations multiples, voire des ambiguïtés lors de travaux ou de projets comportant des équipes pluridisciplinaires. Ainsi le terme « canal » par exemple, ne revêt pas exactement le même sens selon l'approche (marketing, psychologique…). Au sein même d'une discipline comme l'informatique, un canal n'aura pas le même sens pour un spécialiste des réseaux ou pour un expert des interactions hommes-machines. Il nous semble donc pertinent de donner quelques points de repère, afin de fixer la terminologie que nous emploierons dans ce chapitre.

8.3.1.1 Média

Dans la vie courante, un média désigne un support d'information (journal, cassette vidéo, CD audio, DVD, etc.). Dans le domaine des IHM, un média désigne un support technique permettant une communication. Plusieurs points de vue sont étudiés :

– sous l'angle technique, un média est un dispositif physique. Ainsi, un clavier ou une souris sont considérés comme des capteurs, au sens où ils fournissent des informations, en provenance de l'utilisateur et à destination de la machine. Un écran ou une imprimante sont considérés comme des effecteurs, c'est-à-dire qu'ils diffusent de l'information, de l'ordinateur vers l'utilisateur.

– sous l'angle utilisateur, le média peut être perçu comme faisant référence aux éléments sensoriels, perceptuels et cognitifs de l'être humain. On parle alors de média visuel, auditif, audiovisuel, kinesthésique, etc.

8.3.1.2 Multimédia

Un système multimédia est communément identifié comme pouvant supporter plusieurs entrées et/ou plusieurs sorties ; l'élément discriminant un média d'un autre étant la nature de l'information. Un ordinateur multimédia désigne alors un système ou un logiciel informatique qui fusionne plusieurs types de composants tels que le texte, les images, les vidéos... Ce système, capable d'acquérir, de stocker et

---

[3] Adaptation statique des préférences d'usage, des capacités perceptives de l'utilisateur, etc.

[4] Adaptation dynamique, au cours de l'interaction, des préférences d'usage, des capacités perceptives, etc.

de restituer des informations de natures différentes traite les données au niveau élémentaire du signal, sans interprétation sémantique.

Enfin, une distinction peut être faite entre multi-média et multimédia, notamment dans le domaine de l'enseignement, où la première orthographe désigne l'utilisation de plusieurs médias sur plusieurs machines différentes, tandis que la seconde orthographe indique qu'une unique machine contrôle plusieurs médias, grâce à une numérisation des textes, des sons, des images, etc. (cf. chapitre d'Eric Bruillard dans cet ouvrage et [BRU 97]).

8.3.1.3 Mode

Communément, le mode désigne la manière générale dont un phénomène se présente ou dont une action se produit [TRU 00]. Par exemple, « le mode désigne en linguistique la manière dont le verbe exprime l'état ou l'action (impératif, subjonctif, …) » [BEL 92] BELLIK, Y., TEIL, D., Les types de multimodalités. Actes IHM'92. 4èmes Journées sur l'ingénierie des interfaces Homme-Machine, Paris, 30 Nov - 2 Déc 1992.

[BEL 95]. Dans une communication, le mode de communication se rapporte principalement à l'organe (ou au système d'organes) utilisé pour percevoir ou produire des informations. Un mode, dans le domaine de la téléphonie correspond aux normes utilisées par les réseaux. Typiquement, un téléphone devra être bimode pour fonctionner à la fois sur GSM[5] et sur UMTS[6] [WEI 02].

Dans le domaine des IHM, le mode correspond à un état particulier, pour un instant donné, dans lequel se trouve le système interactif. Concrètement, une même action de la part de l'utilisateur sera interprétée de manière différente de la part de la machine, selon le mode en vigueur à un moment donné. Truillet donne l'exemple bien connu de l'éditeur de texte dénommé « vi » du système Unix, qui selon le mode dans lequel il se trouve (mode commande ou mode frappe) n'exécute pas la même action. Un appui sur la touche « x » du clavier permet de taper la lettre « x » de l'alphabet, ou bien cela permet de supprimer le caractère courant, selon le mode dans lequel on se trouve.

Certains équipements sont qualifiés de « multimode ». Ainsi, un boîtier multimode vendu dans le commerce permet aux secrétaires d'utiliser le même micro-casque pour plusieurs tâches au choix (dictée vocale ou conversations téléphoniques). Il comprend un commutateur à trois positions, pour la dictée seule, la conversation seule, ou bien la dictée couplée avec conversation.

---

[5] Global System for Mobile communication : norme européenne pour la téléphonie mobile.

[6] Universal Mobile Telecommunication System : norme européenne destinée à remplacer le GSM et permettant un accès haut débit (2Mbps) aux services d'Internet mobile. Permettra notamment la visiophonie.

8.3.1.4 Modalité

En IHM, la modalité fait référence à la manière d'utiliser les medium. Selon [NIG 96] le terme modalité n'est pas toujours utilisé à bon escient. Pour une même définition donnée, certains appliquent le terme modalité, d'autres le terme média. Globalement, on accepte de dire qu'une modalité est une certaine manière d'utiliser un media. Par exemple, écrire, faire un geste, dessiner sont autant de modalités possibles pour transmettre une information en utilisant le media « stylo ». Comme pour le média, la définition de la modalité dépend du point de vue (utilisateur vs système) que l'on adopte. Truillet par exemple, appréhende la modalité en se plaçant toujours du coté de l'être humain, et en se focalisant sur la structure des informations échangées entre les interacteurs. Il explique qu'une structure telle que le braille peut être perçue aussi bien par le toucher que par le visuel.

Sous l'approche hybride, couplant les deux angles d'étude de la conception d'IHM, l'utilisateur et le système interactif, Nigay et Coutaz proposent *la définition d'une modalité comme le couple <p, r> où :*

p désigne un dispositif physique (par exemple, une souris, une caméra, un écran, un haut-parleur),

r dénote un système représentationnel, c'est-à-dire un système conventionnel structuré de signes assurant une fonction de communication (par ex., un langage pseudo naturel, un graphe, une table).

L'expression suivante fournit une définition complète de la notion de modalité : modalité ::= <p, r> | *<modalité, r>* [NIG 01].

Dans ce contexte, <clavier, langage naturel>, <microphone, langage naturel>, <souris, formulaire graphique> ou encore <ordinateur de poche, langage gestuel> sont des modalités. En désignant une modalité comme étant définie par <modalité, r> on parle également de composition récursive. C'est le cas où une modalité sert uniquement à définir des unités informationnelles pour un autre système représentationnel, le tout formant alors une modalité. Exemple : Pseudo Langue Naturelle écrite ::= < <souris, menu>, Pseudo Langue Naturelle >. La modalité <souris-menu> consiste à sélectionner des bouts de phrases dans un menu. Le tout forme un dispositif logique qui, combiné avec « Pseudo Langue Naturelle », permet d'obtenir la modalité « Pseudo Langue Naturelle écrite ».

8.3.1.5 Multimodal

De manière courante, une installation dite multimodale permet de faire cohabiter des modes qui d'ordinaire sont utilisés séparément. Par exemple, des infrastructures comme la gare de Lisbonne, ou l'aéroport de Lyon St Exupéry,

forment un pôle multimodal, dans le sens où elles accueillent, simultanément des trains, des bus, des métros, des avions, des voitures, etc. Cela offre (en théorie) une efficacité accrue pour l'usager, qui peut bénéficier de la synergie des modes regroupés dans une même infrastructure.

A la différence du multimédia, le concept de multimodal est employé, en informatique, lorsque l'on donne du sens aux informations provenant (ou en direction) de plusieurs médias. Le niveau physique du multimodal fait appel à la perception tandis que le niveau représentationnel fait appel à la cognition. Notons cependant que certains auteurs font référence au multimodal en utilisant le terme multimédia intelligent [NEA 90].

8.3.2 La Multimodalité

Ce terme est important dans le développement de notre chapitre, c'est pourquoi nous lui consacrons une partie importante au sein de ce chapitre. La multimodalité fait référence à l'usage de plusieurs modalités pour la réalisation de la même tâche. Cela doit permettre, en théorie, une *meilleure* interaction et faciliter l'utilisabilité des systèmes proposés. Les critères ergonomiques d'évaluation de cette multimodalité sont souvent le caractère naturel de l'interaction, la robustesse de l'interaction, et sa flexibilité selon le contexte et les types d'usagers.

Depuis les années 80, la multimodalité est étudiée dans le cadre des interactions entre l'homme et la machine. Les chercheurs s'intéressent à la multimodalité dite « en sortie », c'est-à-dire, en variant les canaux de présentation de l'information, en provenance de la machine et à destination de l'utilisateur, mais également, « en entrée », en permettant à l'usager d'interagir de plusieurs manières différentes avec le système, voire de coupler les canaux de communication. Depuis les travaux de Bolt, et son paradigme du « Put that here » [BOL 80], où il combine des entrées vocale et gestuelle, de nombreux travaux ont été menés pour étudier l'utilité et l'utilisation réelle de la multimodalité en IHM.

Les approches théoriques pour décrire et concevoir ces systèmes multimodaux essayent de prendre en compte les visions utilisateur et système dans leurs modélisations. Les usages de la multimodalité, du point de vue de l'utilisateur, sont souvent présentés grâce aux propriétés CARE [COU 94], [COU 95], tandis que les propriétés CASE font référence au point de vue système.

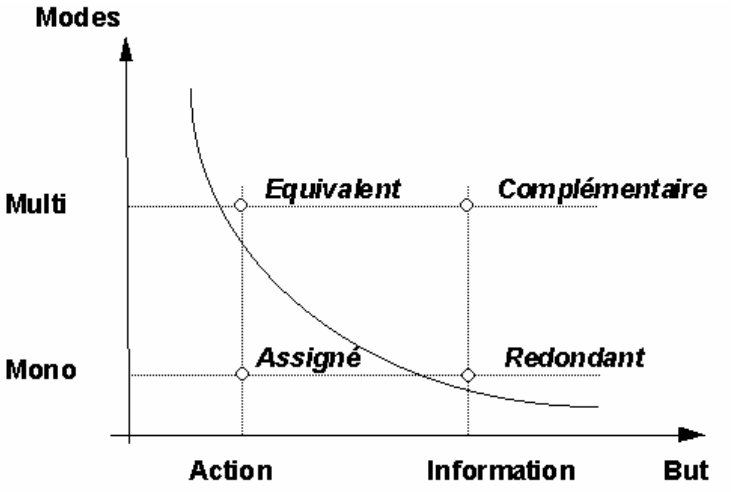

Figure 8.1. CARE : espace de conception centré utilisateur

La Figure 8.1 ci-dessus présente l'espace de conception CARE.

– **C** = Complémentarité. Chaque mode est nécessaire et contribue à la compréhension de l'action.

– **A** = Assignation. L'usager choisit un mode récurrent particulier, ou un sous-ensemble de modes, pour s'exprimer.

– **R** = Redondance. L'usager utilise simultanément plusieurs modes à travers lesquels les informations sont redondantes.

– **E** = Équivalence. L'usager choisit indifféremment tel ou tel mode, ou un sous-ensemble de modes.

La Figure 8.2 ci-après présente l'espace de conception CASE. Jean Caelen explique que cette classification repose sur deux critères : l'usage des médias d'une part qui peut s'effectuer de manière séquentielle ou parallèle, et d'autre part, l'usage des modes (pour l'interprétation ou la génération) qui peut se réaliser de manière indépendante ou combinée [IHM 91].

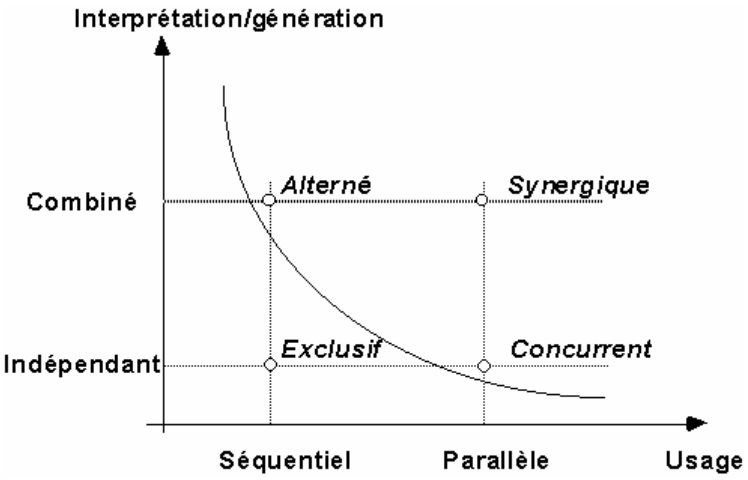

Figure 8.2. CASE : espace de conception centré système

– **C** = Concurrent. Cela fait référence à deux tâches distinctes, en parallèle, sans coréférence. En mode multimodal concurrent, l'usage des médias peut donc se faire de façon parallèle, mais les informations circulant sur ces médias sont indépendantes. Il peut y avoir redondance d'information (exemple : affichage à l'écran et synthèse de la parole) ou conflit si deux commandes contradictoires surviennent en même temps.

– **A** = Alterné. Une tâche est effectuée, avec entrelacement temporel, en coréférence de modalités. En mode multimodal alterné, l'usage des médias est séquentiel et le traitement des informations peut combiner différents modes (exemple : "mets ça là" [BOL 80] en désignant l'objet et le lieu après la fin de la phrase).

– **S** = Synergie. Une tâche est effectuée, en parallèle, en coréférence de modalités. En mode multimodal composé (ou synergique), l'usage des médias s'effectue en parallèle et les traitements sont combinés (exemple : "mets ça là" en parlant et manipulant simultanément). C'est l'aspect le plus proche de la communication naturelle, mais il est complexe à interpréter (cf. ambiguïtés temporelles : dire avant de faire, faire avant de dire).

– **E** = Exclusif. Une tâche unique est accomplie à la fois. En mode multimodal exclusif, deux médias ne peuvent pas être utilisés en même temps. Les informations sont véhiculées par 2 modes indépendants (exemple : clavier, puis souris).

Bellik et Teil [BEL 92] complètent l'approche de Caelen et Coutaz en précisant que les types de multimodalités dépendent de trois paramètres :

– production des énoncés (séquentielle ou parallèle)

– nombre de médias dans un énoncé (un ou plusieurs)

– usage des médias (exclusif ou simultané)

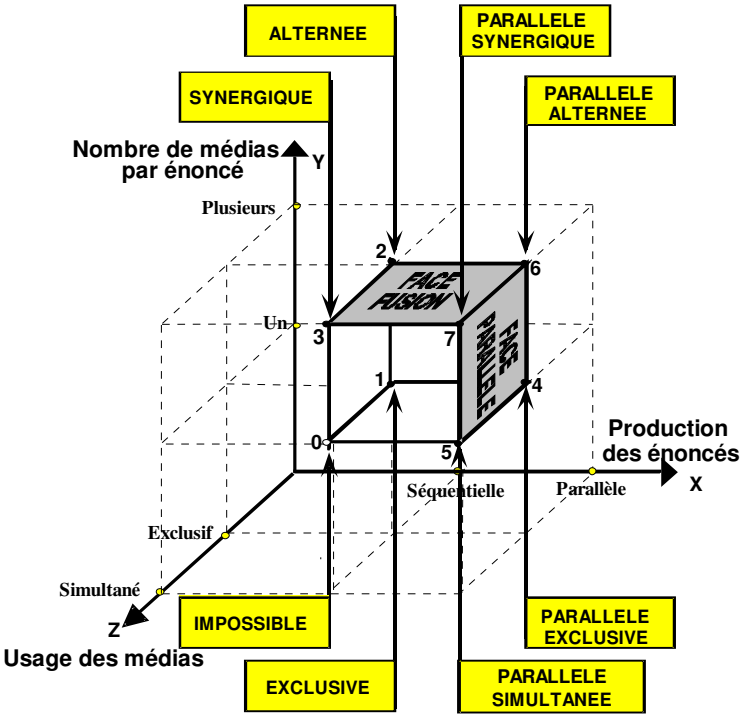

Figure 8.3. Les différents types de multimodalités selon Bellik

Avec cette approche, on a donc sept types de multimodalités possibles ; et non pas huit, puisqu'un sommet du cube représenté en Figure 8.3 est une combinaison impossible : on ne peut pas avoir d'usage simultané des médias si la production des énoncés est séquentielle, avec utilisation d'un seul média par énoncé.

D'autres chercheurs vont encore plus loin. En utilisant les relations de Allen, appliquées aux compositions spatiales, temporelles, articulatoires, syntaxiques et

sémantiques des modalités, ils obtiennent un espace conceptuel représentée par une matrice à 25 positions possibles [VER 00].

Les travaux en multimodalité tendent à montrer que tous les modes de communications ne se valent pas forcément. Différents facteurs seront déterminants pour choisir de concevoir et développer une interface multimodale. Ce choix peut être guidé par la tâche à accomplir : un contrôle aérien, par exemple, pourra éventuellement, être mieux géré si l'humain peut utiliser plusieurs sens simultanément (vision, ouie, toucher). Un chirurgien qui opère un malade, et qui a déjà ses deux mains occupées pourra utiliser d'autres dispositifs pour piloter une caméra endoscopique (pédales, commande vocale). Le type particulier d'utilisateurs visés est bien entendu un facteur déterminant du choix de la modalité (personnes handicapées, personnes âgées, enfants…). Le lieu, le contexte, le type de périphériques disponibles sont encore autant d'éléments devant être pris en compte (milieu bruité/lumineux, utilisation debout/assis).

D'autres pistes de recherche suggèrent l'emploi de la multimodalité pour pallier les problèmes rencontrés par les interactions traditionnelles basées sur la manipulation directe et le pointage. Cela est valable lorsque l'utilisateur a besoin d'expliciter ses besoins, lors d'une recherche d'information, par exemple, et que les systèmes classiques ne sont plus d'aucune utilité [ROU 01]. Ce sera certainement le cas également, lorsqu'il y aura négociation entre des agents, dans le cadre de l'enseignement à distance ou du commerce électronique. On fait alors référence aux 5 sens humains : montrer le produit, le faire toucher (par systèmes à retour d'effort), le faire sentir (parfums de synthèse proposés par les 3 Suisses à des œnologues, ou expériences dans des cinémas allemands), faire entendre (le claquement d'une portière, gage de fiabilité et de sécurité), etc. Cependant, le fait de concevoir une application de manière multimodale ne présage en rien de la bonne utilisation (et utilisabilité) de cette application. En effet, il est illusoire de croire qu'un système, parce qu'il est multimodal, sera forcément utilisé de manière multimodale par les utilisateurs [OVI 99]. Sharon Oviatt présente d'ailleurs à ce propos les dix mythes de la multimodalité :

– 1) ce n'est pas parce qu'une interface est multimodale que les utilisateurs vont utiliser la multimodalité.

– 2) le pattern parole-pointage n'est pas le plus intéressant.

– 3) la multimodalité ne signifie pas obligatoirement « parallélisme ».

– 4) la parole n'est pas un mode « de base » dans un système multimodal.

– 5) le langage multimodal ne diffère pas du langage unimodal.

– 6) l'interaction multimodale ne favorise pas la redondance.

– 7) les erreurs sur un mode ne sont pas compensées par un autre mode.

– 8) les utilisateurs n'organisent pas « leur » multimodalité de la même manière.

– 9) les modes ne sont pas équivalents.

– 10) un système multimodal n'est pas plus efficace qu'un autre.

8.3.3 Dialogue et multimodalité

La multimodalité couplée au dialogue en langage naturel demeure une piste de recherche intéressante, et donne lieu à de nombreux travaux. Plus encore, des interactions multimodales en relation directe avec des activités de communications humaines sont offertes dans bon nombre d'outils que nous utilisons quotidiennement, mais nous avons peu conscience qu'elles le sont effectivement. Nous nous focalisons sur quelques domaines particuliers comme les jeux vidéo (cf. manettes vibrantes, volants et pédales pour simulateurs), ou quelques logiciels de bureautique[7], mais nous ne prenons pas conscience immédiatement que bon nombre de nos interactions courantes avec les machines qui nous entourent sont multimodales. Quelques exemples suffiront à étayer notre propos :
(a) l'ascenseur qui annonce vocalement des informations (« vous arrivez à tel étage » ou bien, « attention, vous êtes devant la cellule » (sous-entendu vous empêchez la porte d'ascenseur de se refermer) ;
(b) le signal sonore de votre véhicule qui indique que vous avez oublié d'éteindre vos phares lorsque vous ouvrez votre portière et que le contact est coupé ; dans les véhicules haut de gamme, tel le Xsara Picasso de Citroën, ce signal sonore et couplé à l'affichage d'un message digital sur le tableau de bord, en langue naturelle, du type « phares allumés » ;
(c) le téléphone portable qui permet d'attribuer à tel appellant une sonnerie particulière : cognitivement, on obtient plus d'informations car le sens de la sonnerie n'est plus simplement « vous avez un appel » mais plus encore « vous avez un appel de la part de telle personne » ; libre à vous de décrocher ou pas compte tenu de la valeur ajoutée de cette information.
(d) panneau digital et lumineux comme le plan de la ville dans une gare, que l'on peut toucher — en entrée — et qui allume des diodes pour former un itinéraire et qui parle — en sortie — (cf. Figure 8.4) ;
(e) son 3D dans les jeux vidéo, palonnier, volant vibrant, manettes à retour d'effort, tapis de combat (cf. Fighting Arena sur la Figure 8.5).

---

[7] Le pédalier-secrétariat proposé par l'entreprise Dragon permet de piloter au pied la réécoute des textes transcrits dans Dragon NaturallySpeaking. Il propose trois fonctions : écoute à partir de la positon du curseur (pédale centrale), déplacement au paragraphe précédent (pédale de gauche), déplacement au paragraphe suivant (pédale de droite).

Tous ces exemples sont autant de mise en œuvre de dispositif qui peuvent être considérés comme multimodaux, dans le sens où ils autorisent l'utilisateur à choisir un moyen particulier (parmi plusieurs possibles) de communiquer avec son environnement. Ce moyen d'interagir avec un dispositif n'est pas forcément le meilleur, mais il est une alternative possible.

Figure 8.4. Plan digital tactile et lumineux

Figure 8.5. Dispositif pour coups de pied et de poing pour console de jeu Playstation

Nous avions proposé dans [ROU 00] un dialogue multimodal sur Internet permettant d'effectuer une recherche d'information documentaire de plusieurs manières possibles : par manipulation classique (clavier/souris), par interaction vocale en langue naturelle, ou par couplage des deux. De même en sortie, nous avions étudié les apports d'une interaction vocale dialoguée (avec synthèse vocale calculée à la volée) et présentation graphique des résultats en vue dite « fish-eye-

view » (vue en œil de poisson). Notre constat était un manque de langage des programmations pouvant permettre la réalisation de dispositifs multimodaux. De tels langages de programmation commencent aujourd'hui à voir le jour. Le W3C a proposé par exemple EMMA[8] (Extensible MultiModal Annotation language - http://www.w3.org/TR/EMMAreqs) et Ink Markup Language[9] (http://www.w3.org/TR/inkreqs) qui prennent en compte la mise en œuvre et la gestion d'interfaces multimodales sur le réseau Internet.

8.4 Mise en œuvre

L'utilisation de la multimodalité en IHM permet, entre autres, de donner à l'utilisateur plus de liberté quant à la manière d'interagir avec un système informatique. Le fonctionnement du système demeure le plus souvent inchangé, mais la manière de récolter (respectivement « de restituer ») les données en provenance (respectivement « à destination ») de l'utilisateur diffère selon les contextes d'utilisation, les périphériques disponibles, les profils d'usagers identifiés, etc. Dans les lignes qui suivent, nous présentons des exemples de systèmes informatiques multimodaux. Bien qu'il soit possible d'utiliser simultanément de la multimodalité en entrée et en sortie, nous exposons volontairement les deux cas séparément, pour faire apparaître les apports de chacun.

8.4.1 Multimodalité en entrée

Prenons le cas d'un logiciel consacré à la recherche d'informations cinématographiques. L'interface graphique d'une application permettant de sélectionner des films en fonction de leur durée pourrait ressembler à celle proposée en Figure 8.6. Sur un ordinateur personnel, il est aisé de saisir ces données au travers d'un formulaire.

---

[8] EMMA is a target data format for the semantic interpretation specification being developed in the Voice Browser Activity, and which describes annotations to speech grammars for extracting application specific data as a result of speech recognition. EMMA supersedes earlier work on the natural language semantics markup language in the Voice Browser Activity.

[9] The Ink Markup Language will serve as the data format for representing ink entered with an electronic pen or stylus in a multimodal system. The markup will allow for the input and processing of handwriting, gestures, sketches, music and other notational languages in web-based multimodal applications.

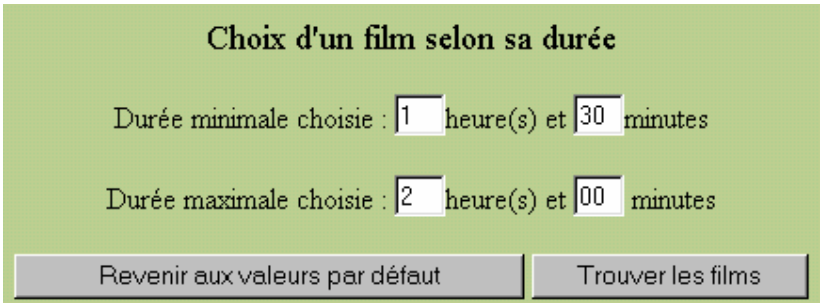

Figure 8.6. Interface graphique permettant de sélectionner un film selon sa durée

Sur la majorité des périphériques actuels et dans un proche avenir, on peut envisager une saisie vocale en langue naturelle. Avec une reconnaissance vocale on aurait par exemple un dialogue tel que celui de la Figure 8.7.

(…)
Homme : Je veux effectuer une sélection selon la durée du film
Machine : Quelle est la durée minimale souhaitée pour ce film ?
Homme : Pardon ?
Machine : Vous devez préciser la durée minimale des films recherchés (par exemple 1 heure 40 minutes)
Homme : une heure trente
Machine : Quelle est la durée maximale souhaitée pour ce film ?
Homme : deux heures
Machine : J'ai trouvé 16 films ayant une durée comprise entre 1h30 et 2h. Souhaitez-vous connaître les titres de ces films ?
(…)

Figure 8.7. Exemple de dialogue oral permettant de sélectionner un film selon sa durée

Ici, la multimodalité ne change rien au service rendu par le système informatique, puisque dans un cas comme dans l'autre, seize films répondant à la requête de l'utilisateur ont été trouvés. Idéalement, on doit être en mesure d'utiliser indifféremment l'un ou l'autre des moyens d'interaction proposés, sans surcharge cognitive supplémentaire ou besoin d'apprentissage particulier. En pratique, bien entendu, chaque modalité à un coût propre, qui n'est pas forcément équivalent à celui des autres modalités disponibles. De même, l'utilisation conjointe de modalités ne garantie en rien une meilleure efficacité.

8.4.2 Multimodalité en sortie

Le cas d'étude que nous présentons ci-après traite de l'utilisation de la multimodalité dans le cadre d'une interaction dialoguée (pour partie). Une situation de jeu est mise en œuvre. Le support utilisé est le réseau Internet et la technologie MSA (Microsoft Agent). Microsoft propose, en effet, des agents animés que l'on intègre aux pages Web, en insérant du code dans l'entête d'une page HTML. Ceci est censé améliorer l'interaction entre l'homme et la machine, puisque l'agent peut adopter des attitudes explicites (comme la surprise, l'explication, l'incompréhension, ….), se déplacer sur l'écran, désigner des parties de l'écran (à gauche, à droite, etc.). Ces MS-Agent (http://msdn.microsoft.com/msagent) disposent de logiciels de synthèse et de reconnaissance vocale (en anglais, en français, etc.), qui s'installent sur la machine cliente, lors du chargement de la page Web.

Dans notre étude, purement descriptive, il s'agit pour la machine de tirer un nombre au hasard entre deux bornes (par défaut entre 0 et 100). L'utilisateur doit entrer une proposition numérique dans une boîte de dialogue et valider cette proposition en cliquant sur un bouton. Bien que nous nous placions dans une situation « d'apprentissage », dans laquelle l'utilisateur (un enfant par exemple) pourra éventuellement découvrir le mécanisme de dichotomie, à travers le jeu, et à l'aide d'une application multimodale, nous présentons ce dispositif informatique uniquement à titre d'illustration, en ne nous intéressant qu'aux possibilités techniques qu'il apporte. Nous ne traitons en aucun cas du contenu pédagogique du logiciel, et ne nous prononçons pas sur le bien fondé de cette démarche sur un plan éducatif.

Deux profils ont été programmés : novice et expert. En mode expert, le système se contente de répondre « c'est plus que X » ou bien « c'est moins que X » (X mis pour la valeur entrée par l'utilisateur). Dans ce mode, Merlin ne fait pas de gestes. On présente des informations textuelles à l'écran, et l'on diffuse, en parallèle, une voix de synthèse qui prononce exactement la même chose que ce qui a été affiché. En mode novice, l'utilisateur est aidé par l'agent animé, qui lui rappelle les bornes en vigueur, et qui fait des gestes pour expliciter ces bornes.

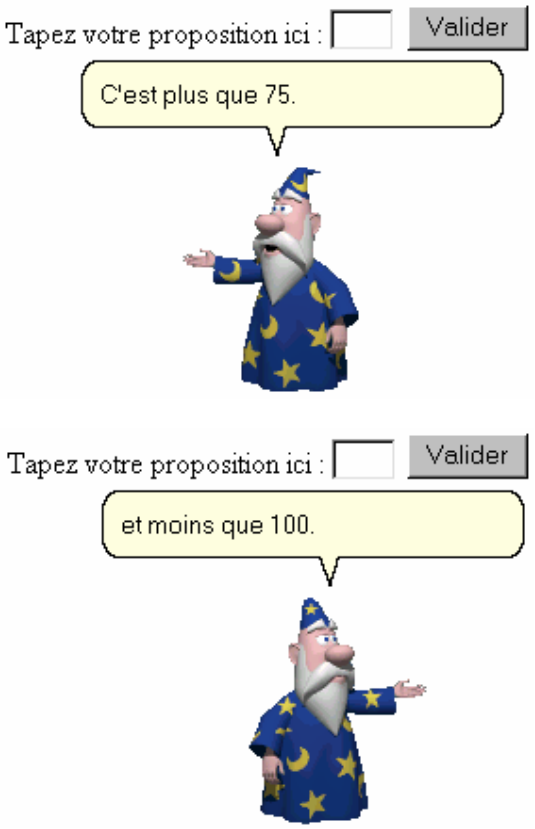

Figure 8.8. L'agent animé Merlin utilise des gestes pour aider l'utilisateur

Dans l'exemple de la Figure 8.8, nous observons une interaction entre les interlocuteurs. L'utilisateur novice vient d'entrer la valeur 75. L'ordinateur compare la valeur donnée avec la valeur à trouver, et calcule les éléments à présenter en réponse. Nous avons programmé trois types d'informations simultanées données à l'usager par l'agent animé Merlin. En effet, il énonce oralement, grâce à une voix de synthèse « c'est plus que 75 et moins que 100 ». D'autre part, ces informations sont également affichées sur l'écran, dans des bulles de dialogue (dont on peut éventuellement paramétrer la taille de la police de caractères). Enfin, Merlin accompagne ses énoncés de gestes de la main, orientés vers la gauche (de l'écran), pour indiquer la borne inférieure, puis vers la droite, pour donner la borne supérieure. Notons que si ces informations sont considérées

comme redondantes, ici, dans le cadre précis de cet exercice, nous ne considérons pas qu'elles sont strictement équivalentes, d'un point de vue sémantique, étant donné qu'un découpage de l'énoncé est effectué, pour distinguer la borne inférieure de la borne supérieure.

Dans l'exemple de la Figure 8.9, l'utilisateur a demandé de l'aide au système. Merlin lui rappelle la valeur des bornes de la session en cours (80 et 50) et explique comment procéder par dichotomie pour proposer une valeur à tester optimale (65 dans notre exemple, c'est-à-dire 50+15 ou 80-15).

Figure 8.9. Merlin explique comment procéder à une dichotomie

L'usager est aidé par le fait qu'il ne doit pas forcément mémoriser les bornes, puisqu'elles sont recalculées et redonnées à chaque intervention de l'agent animé. D'autre part, on entend et voit simultanément l'explication qu'il donne.

Un apport important de cette multimodalité en sortie semble être la possibilité de fournir des éléments pertinents pour une meilleure compréhension, de la part de l'utilisateur, de ce que fait l'ordinateur. Les animations (Merlin qui écrit dans un livre, qui se frotte le menton, qui pense, etc.) sont autant de points permettant de comprendre ce que fait le système à chaque instant (cf. théorie de l'activité de Donald Norman [NOR 86]). L'agent animé est alors perçu, semble-t-il, comme la machine elle-même ; l'utilisateur fusionnant mentalement les capacités de la machine à celles de son interlocuteur. Cela reste cependant à vérifier de manière scientifique, en procédant à des mesures et des évaluations (en laboratoire et en situation réelle d'utilisation) qui jusqu'à présent n'ont pas encore été effectuées.

8.5 La multimodalité au cœur des IHM

Nous avons procédé à un rappel des notions importantes utilisées en communication homme-machine dans le cadre de la multimodalité. La souplesse et le naturel escomptés laissent à penser qu'à terme, toute interface sera utilisable de manière multimodale. C'est déjà le cas pour bon nombre de tâches quotidiennes mais nous n'y portons pas forcément attention. Hormis certains cas particuliers, pour lesquels des contraintes sont fortes (handicap de l'utilisateur), les apports de la multimodalité au sens des IHM résident, à notre sens, dans la liberté de choix laissée à l'utilisateur lors de la manipulation de l'interface. Une des critiques classiques, à l'égard de la multimodalité, que l'on entend parfois de la part de personnes qui ne sont pas spécialistes du domaine, consiste à s'interroger sur le bien fondé d'une telle approche. Le néophyte tente maladroitement de montrer que l'approche multimodale n'est pas convaincante, sous prétexte qu'une approche dite « traditionnelle » peut en faire autant.

L'apport de la multimodalité est tout autre. Cela permet à l'utilisateur de partager du sens avec la machine par plusieurs moyens d'interactions différents (complémentaires, redondantes, mutuellement exclusives, etc.). Chaque utilisateur choisit le meilleur type d'interaction possible selon certains critères : la tâche à accomplir, le contexte, les habitudes, l'environnement (bruité, lumineux …), le niveau de stress, le temps imparti pour mener à bien cette tâche, etc.

Actuellement, les travaux en multimodalité évoluent vers les modalités dites sensorielles, de présentation, et d'action. On s'intéresse de plus prêt aux interactions en grand (salle où l'utilisateur se trouve immergé, tableau « magique » avec lequel il peut interagir dans un environnement de réalité augmentée [COU 02]) et en petit (dispositifs miniaturisés, dispositifs mobiles, sans fil, etc.). Dans le cadre de leurs recherches sur les gestes médicochirurgicaux assistés par ordinateur,

des chercheurs grenoblois du laboratoire TIMC[10] ont mis au point un système multimodal, permettant à un chirurgien d'utiliser sa langue comme transmetteur[11] d'information. Lors de la manipulation d'un instrument dans le corps du patient, le médecin reçoit à distance de très faibles impulsions électriques qui lui indiquent toute déviation de sa trajectoire. Il peut ainsi corriger son geste, sans avoir à consulter les traditionnels moniteurs (écrans de contrôle). Ce sont donc d'autres informations que celles conventionnellement utilisées qui font sens ici.

L'exemple que nous avons présenté utilisait la technologie MSA pour illustrer notre propos, et faire apparaître les apports de la multimodalité, tant en entrée qu'en sortie. D'autres recherches dans cette voie, utilisant des agents animés, sont engagées au MIT par exemple, dans lesquelles le contexte joue une rôle important pour la compréhension des situations. Le « context aware assistant » de [YAN 00] est en effet un agent animé capable de gérer une situation de filtrage à l'entrée d'un bureau par exemple, en utilisant des capteurs sensoriels qui lui permettent de savoir combien de personnes sont déjà présentes dans la pièce. Il cherche ensuite la meilleure façon de gérer l'activité en fonction des contraintes : faire patienter le visiteur, lui proposer un autre rendez-vous selon les disponibilités de chaque personne (visiteur / visité), envoyer un message électronique à la personne travaillant dans son bureau pour lui signifier une urgence, etc.

Par ailleurs, on observe que certaines nouvelles interfaces personnes/organisation sont disposées à intégrer des composants multimodaux dans leurs architectures. C'est le cas notamment pour les grands groupes de vente à distance, qui, par l'intermédiaire de systèmes vocaux interactifs gérés dynamiquement, pourraient résoudre leurs problèmes de maintenance de messages oraux préenregistrés. On voit également apparaître des travaux d'étude sur le couplage entre les notions de multicanal (utiliser plusieurs canaux : e-mail, fax, SMS, etc.) et de multimodalité, notamment dans les domaines du commerce électronique [DER 02], de l'apprentissage et de l'enseignement à distance. Enfin, la personnalisation va croître, nous semble-t-il, pour répondre de manière beaucoup plus appropriée aux attentes des utilisateurs. Encore une fois, les avancées significatives ne peuvent pas se faire dans un seul champ d'étude. Si les technologies informatiques ont bien entendu un rôle primordial à jouer, le couplage avec les domaines connexes de la cognition devraient créer des conditions favorables pour faire émerger des nouveaux outils informatiques utiles, puissants, robustes et malléables.

---

10 http://www-timc.imag.fr/gmcao/index.html

11 La langue transmet les informations au cerveau. C'est une multimodalité en sortie : la langue joue le rôle de capteur et non pas d'effecteur, dans ce cas précis.

Bibliographie

[BEL 92] BELLIK, Y., TEIL, D., Les types de multimodalités. Actes IHM'92. 4èmes Journées sur l'ingénierie des interfaces Homme-Machine, Paris, 30 Nov - 2 Déc 1992.

[BEL 95] Y. BELLIK, D. BURGER, The Potential of Multimodal Interfaces for the Blind: an Exploratory Study, RESNA'95, Vancouver, Canada, 9-14 June 1995.

[BOL 80] BOLT, R.A., Put-that-here: voice and gesture at the graphic interface. Computer Graphics, 14, 262-270, 1980.

[BRU 97] BRUILLARD, E., Les machines à enseigner, Éditions Hermès. Paris, 1997.

[CHA 02] CHARPAK, G., BROCH, H., *Devenez sorciers devenez savants*, éditions Odile Jacob, Paris, 2002.

[COU 94] COUTAZ, J., NIGAY, L., Les propriétés "CARE" dans les interfaces multimodales, IHM'94, Lille 1994.

[COU 02] COUTAZ, J. LACHENAL, C. BERARD, F. BARRALON, N., Quand les surfaces deviennent interactives, in Les cahiers du numérique, Lavoisier, Vol. 3, Numéro 4-2002, pp.101-126.

[COU 95] COUTAZ, J., NIGAY, L., SALBER, D., BLANDFORD, A., MAY, J., Young R., Four Easy Pieces for Assessing The Usability of Multimodal Interaction: The CARE properties, InterAct'95, Lillehammer (Norway), June 1995, pp. 115-120.

[DER 02] DERYCKE, A., ROUILLARD, J., La Personnalisation de l'Interaction dans des Contextes Multimodaux et Multicanaux : une Première Approche pour le Commerce Electronique, IHM 2002, Poitiers, 2002.

[EFR 41] EFRON, D., Gesture, Race and Culture. The Hague: Mouton & Company, 1972. Reprinted from Gesture and Environment, New York: King's Crown Press, 1941.

[EKM 69] EKMAN, P., FRIESEN, W., The repertoire of Non-verbal behavior: Categories, Origins, Usage, and Coding. Semiotica, 1, 49-98, 1969.

[EKM 75] EKMAN, P., FRIESEN, W., Unmasking the face. New Jersey: Prentice-Hall, 1975.

[EKM 78] EKMAN, P., FRIESEN, W., Facial action coding system. Palo Alto, CA: Consulting Psychologists Press, 1978.

[IFR 94] IFRAH, G., Histoire universelle des chiffres, Robert Laffont, Tome 1 et 2, Paris, 1994.

[IHM 91] Production des participants en ateliers, IHM'91, Troisièmes Journées sur l'Ingénierie des Interfaces Homme-Machine, Dourdan, 11-13, 1991.

[KEI 01] KEIM, D., Visual Exploration of large data Sets. Communications of the ACM, Vol. 44., N. 8, 2001, p. 39-44.

[LEM 86] LE MOIGNE, J.L., Genèse de quelques nouvelles sciences : de l'intelligence artificielle aux sciences de la cognition, Intelligences des mécanismes - Mécanismes de l'intelligence, Fayard, 1986.

[MCN 92] MC NEILL, D., Hand and Mind: What gestures reveal about thought, Chicago, IL, University of Chicago Press, 1992.

[NEA 90] Neal, J. G., Shapiro, S. C., Intelligent Multi-Media Interface Technology, in Intelligent User Interfaces, J. W. Sullivan and S. W. Tyler, eds., ACM Press, 1990.

[NIG 96] NIGAY, L., COUTAZ, J., *Espaces de conception des interfaces multimédia et multimodales.* Revue : TSI, numéro spécial Multimédia et Collecticiel, Volume 15, N° 9, 1996, AFCET & HERMES Publ, pp. 1195-1225.

[NIG 01] NIGAY, L., Modalité d'Interaction et Multimodalité, Habilitation à Diriger des Recherches, spécialité Informatique de l'Université Joseph Fourier - Grenoble I, 2001.

[NOR 86] NORMAN, D., User Centered System Design, New Perspectives on Human-Computer Interaction, Lawrence Erlbaum Associates, Hillsdale, NJ, 1986.

[OVI 99] OVIATT, S., Ten myths of multimodal interaction. Communications of the ACM, Vol. 42, N. 11, 1999, p. 74-81.

[PEK 02] PEKKOLA S., HIEKKILÄ J., TUUNAINEN V., Launching Multi-Modal Interaction on Ec-site. Proceedings of the 35th Hawaii International Conference on System Sciences, IEEE press, 2002.

[POY 80] POYATOS, F., Interactive functions and limitations of verbal and Nonverbal behaviors in natural conversation, Semiotica, 30-3/4, 211-244, 1980.

[PRI 96] PRINCE, V., Vers une informatique cognitive dans les organisations, Collection Sciences cognitives, Masson, 1996.

[RIM 91] RIMÉ, B., SCHIARATURA, L., Gesture and speech. In R.S. Feldman & B. Rimé, Fundamentals of Nonverbal behavior, 239-281. New-York : Press syndicate of university of Cambridge, 1991.

[ROU 00] ROUILLARD, J., Hyperdialogue sur Internet. Le système HALPIN, Thèse de doctorat d'informatique, Université Grenoble I, 2000.

[ROU 01] ROUILLARD, J., Dialogue et Multimodalité. Revue RIHM - Revue d'Interaction Homme-Machine. Vol 2, N°1, pp.99-125, 2001.

[ROU 02] ROUILLARD, J., A multimodal E-commerce application coupling HTML and VoiceXML, Eleventh International World Wide Web Conference, Waikiki Beach, Honolulu, Hawaii, USA, 2002.

[THO 96] THORISSON, K. R., Communicative Humanoids A Computational Model of Psychosocial Dialogue Skills, Ph D, MIT, 1996.

[VER 00] VERNIER, F. NIGAY, L., Espace de Conception pour les Interfaces Multimodales, Acte du colloque sur les interfaces multimodales, Grenoble, 2000.

[TRU 00] TRUILLET, P., Support de cours d'IHM de DESS SIGMA, Université de Toulouse, 2000.

[WAL 98] WALTER, H., *Le français d'ici, de là, de là-bas,* 1998.

[WEI 02] WEISS, S., Handheld Usability, John Wiley & Sons, 2002.

[YAN 00] YAN, H., SELKER, T., *Context-Aware Office Assistant*, proceedings of 2000 International Conference on Intelligent User Interfaces, New Orleans, Louisiana, 2000.

Index

Table des matières